DOCUMENT RESUME

ED 030 944 EA 002 260

By-Finn, Jeremy D.

Evaluation of Instructional Outcomes: The Use of Unstructured Data.

Pub Date 6 Feb 69

Note-21p.; Preliminary Draft of paper presented at the Annual Meeting of the Amer. Educ. Res. Assn. (Los Angeles, Calif., Feb. 6, 1969).

EDRS Price MF-$0.25 HC-$1.15

Descriptors-*Behavior Rating Scales, Data Analysis, Data Collection, *Educational Objectives, *Evaluation Methods, *Measurement Techniques, *Models, Primary Grades

Encouraging a broader view of educational evaluation by emphasizing systematic means for the collection and analysis of a wide variety of types of evaluation data at all educational levels, an extension is proposed to the evaluation rationale formulated several years ago by Ralph W. Tyler. The revised model, patterned after the "multiform organismic system of assessment" devised by the Office of Strategic Services, develops a principle of multiple measurements, placing basic data (test scores, teacher ratings, etc.) in a holistic spectrum consisting of many sorts of objective and subjective data. A "behavior dictionary" of measurable indicators of educational achievement is being developed as a tool for the analysis of quantitative evidence. The model, measuring cognitive, affective, and habitual behaviors in class and school situations, is being developed for initial application at the elementary school level. (JK)

ED030944

EVALUATION OF INSTRUCTIONAL OUTCOMES: THE USE OF UNSTRUCTURED DATA*

Jeremy D. Finn

Faculty of Educational Studies

State University of New York at Buffalo

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

January 31, 1969

PRELIMINARY DRAFT

NOT TO BE CITED OR QUOTED

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY Jeremy D. Finn

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

EA 002 260

* Paper presented at the 1969 Annual Meeting of the American Educational Research Association, Los Angeles, California, February 6, 1969.

Evaluation of Instructional Outcomes: The Use of Unstructured Data

Jeremy D. Finn

Faculty of Educational Studies

State University of New York at Buffalo

About forty years ago, Ralph Tyler formulated a systematic methodology of evaluation of the outcomes of instruction. To this day, Tyler's evaluation model has remained a predominant influence upon evaluation theory, surviving both a variety of interpretations for actual classroom practice, and attacks upon its adequacy. The reasons for its sustainance are perhaps obvious but nvertheless important. For one, the model includes the more common evaluation practices actually employed by teachers in the classroom setting. These practices, which may be termed measurement rather than evaluation, involve the construction and administration of unit or semester tests, and the assigning of course grades based on the test results. Second, the Tyler model encompasses a number of practices educators would like to apply, but for which they probably don't have the time. These include the evaluation of change that has taken place during a course rather than simply of the final status of the students. Also, full use of diagnostic evaluation data obtained during a course to point to specific weaknesses in individual learning, or to instruction in particular areas is rarely made. The data are often not employed for subsequent modification of the course methods and objectives to compensate for the weaknesses, i.e. "formative" evaluation. The Tyler evaluation model also stresses the need to consider a wide variety of instructional outcomes. While teachers' "statements of objectives" usually acknowledge this need, their formal evaluations frequently do not, and tend to rest primarily on the results of a very small number of paper-and-pencil tests. Further, Philip Jackson has observed that "...teachers treat (testing) as being of minor importance in helping

them understand how well they have done."[1] Thus even the limited amounts of data collected are often not fully exploited.

I do not mean to suggest that the principles outlined by Tyler are not followed at all, however. Many primary grade report cards include data on a variety of achievements, as "makes an effort to succeed, uses time to best advantage, attempts to solve own problems." While these may be important achievements, and while teacher judgements on these variables certainly constitute important evaluations, it would seem that such evaluations lack in both objectivity and validity. Evaluatory data are employed in a "formative" manner in a number of better schools. Students are grouped for reading and arithmetic once some indication of their ability has been obtained. In some schools, second grade pupils are allowed to continue where they "left off" in their first grade readers. This is not done frequently, however, and is rarely done at early ages in subjects as art, music, social studies, science or gymnastics. The primary reason for this is the simple but convincing lack of instructional staff. In addition, there is a notable paucity of evaluation methodology for use in these areas, especially with young children. This lack tends to make the task even more formidable.

It is the purpose of this paper to propose an extension to the Tyler evaluation rationale which encourages a broader view of educational "evaluation" by emphasizing systematic means for the collection and consideration of a wide variety of types of evaluation data, at all educational levels. The modifications are proposed to facilitate the satisfying of the basic principles of the original model, in actual practice. The revised model gives consideration to the basic data with which we are familiar (i.e. test scores, teacher ratings), but places them in a broader spectrum consisting of many sorts of objective and subjective data.

There are currently changes in thinking about educational policy and practice which emphasize the need for an expanded evaluation

model. One of these concerns evaluation practices themselves. Education specialists, reacting to current and predicted social pressures, are beginning to emphasize the need for evaluation of comparatively large units of instruction, such as sequences of courses in a given area, entire grade levels or several grades, or even of entire schooling careers. This need is resulting in the conduct of global evaluation projects, as the State of Pennsylvania Educational Quality Assessment, the New York Quality Measurement Project, the National Assessment Project, and in the conduct of such studies as Equality of Educational Opportunity and Project Talent. The aspect of these studies of greatest significance to the evaluator is the interest shown in objectives of the education process not typically evaluated in any particular classroom. For example, the State of Pennsylvania evaluators have listed these ten "goals" of the public educational system.[2] The student is expected to acquire:

1. Self understanding and self acceptance.
2. Understanding and appreciation of social, cultural, and ethnic groups different from his own.
3. Mastery of the use of words and numbers.
4. Positive attitudes toward school and learning.
5. The habits and values associated with responsible citizenship.
6. Good health habits and knowledge of the conditions necessary for maintaining physical and emotional well-being.
7. Experience with creative processes in a variety of fields.
8. A full grasp of the opportunities for preparing for a productive life.
9. Understanding and appreciation of the natural sciences, the social sciences, the humanities, and the arts.
10. Preparation for effective participation in a changing world.

While we may wish to take issue with one or more of these objectives, the fact remains that there are a number of outcomes of the entire education of the individual deemed important, beyond a given level of cognitive attainment. Largely, these

"other" objectives are concerned with preparation of the individual for his integration into an adult society (see e.g. objectives 1, 2, 5, 6, 8, 10). And for the most part, these "other" objectives are not formal goals of any particular course or course unit.

Second, we are currently witnessing a large increase in responsibility placed upon the educational sector of our society for the solution of major cultural problems, such as those of cultural deprivation, poverty, and job retraining. Our evaluation needs for related educational programs are different from any we have experienced before. There is both the need to consider a much broader range of instructional objectives (e.g. "to increase the 'self-image' of the Negro child; to learn methods and places for obtaining employment"), and the need to consider "special groups" of students who may not be amenable to evaluation through the more traditional paper-and-pencil techniques (e.g. pre-school or primary grade children, the illiterate, the bilingual).

Finally, we are well into a period of time during which new curricula are regularly being introduced into the classroom. With each "innovation" comes a new set of objectives, often quite different from those of the older programs. Many of the new mathmatics curricula, for example, are directed toward increasing student "interest" in the area, with the expectation that as a result, knowledge and understanding of the subject matter will ultimately increase. And certainly we see new types of outcomes accompanying programs in individual or computerized instruction.

I will summarize the implications of these trends for an expanded evaluation paradigm. Modifications in the basic model must present specific principles and procedures for evaluating a wide variety of instructional outcomes; both the cognitive and the non-cognitive; in order that a "total picture" of results may be arrived at for either small or large educational units. In allowing for this variety in outcomes, the expanded model should allow for consideration of

a variety of types of data, which can be combined in such a manner as to yield valid indicators of the extent to which a number of educational goals are being met, and the extent to which they are not. That the usual evaluation procedures do not meet these needs has recently been indicated by Tyler

> ...it was recognized that present tests involve some limitations in their usefulness both for appraising particular developments and also for giving more general kinds of knowledge. Present tests are commonly limited to only a few of the important objectives emphasized in different... programs; and, for this reason, do not provide all appraisal information needed. Also, most of the available tests have been developed for measuring class or school central tendencies, and/or individual differences among pupils. They rarely sample adequately the range of achievement of pupils in the several school subjects. Hence, one has little basis for judging the progress made by the lower part or the upper part of the class distribution, yet researchers often wish to identify which students are making progress and which have difficulty. Furthermore, in trying to appraise adequately particular instructional materials or devices, it is usually necessary to sample the expected learning rather comprehensively for the particular aspects that the device was developed to facilitate. Most current tests are not providing such samples. There seemed to be general agreement that the development of more adequate appraisal procedures and materials was a part of the needed research for...educators to undertake.[3]

The principles of data collection and analysis needed for the expanded evaluation paradigm have already been developed for use in another context, but have not been employed in educational settings. The principles to which I refer are those of "assessment," as the term was employed by the OSS Assessment Staff during the 1940's, and are outlined in detail in Assessment of Men.[4,5] The Assessment project involved a series of short, extremely intensive periods during which candidates for jobs as espionage agents were subjected to a wide variety of tests and testing situations. The latter included a large number of simulated action situations. Although the length of time for the evaluation of any one group (class) of subjects was measured in days and weeks, long periods of preparation were needed. The assessment procedure was termed the "multiform organismic system of assessment," "multiform" indicating essentially that multiple measures were collected to provide evidence on the behaviors of interest, and "organismic" indicating that a number of measures were

combined in some fashion to provide evidence on more global traits than would be tapped by any single scale. These principles of course are the very ones we would like to employ in our evaluation model.

"Assessment" is basically an eight-step procedure, beginning with a preparatory analysis of all the jobs for which the candidates are to be assessed. For us, the "jobs" are defined by the objectives to which we expect the course or school to address itself. The second step is to describe in all ways possible the behavior of individuals who are or who are not successful in performing the jobs. Third, assessment involves developing quantitative indices of each of the indicators of success, and fourth, tests and natural situations are designed which give the assessees the opportunity and compulsion to display behaviors of importance. For us, the classroom situation is one very important natural situation to consider. The fifth step of the assessment involves the combining of the numerous "bits" of evidence gathered at the previous stage to yield a holistic picture of the degree to which the jobs are being fulfilled (i.e. the degree to which the objectives are being met.) The final stages of assessment involve preparing non-technical summaries of the data collected, holding group conferences for reviewing and correcting the data summaries, and preparing methods for appraising the extent to which the assessment procedures have been successful. While the final three stages are important aspects of educational evaluations, I will focus on the first steps which are the most relevant here.

It may seem at first glance that the steps outlined present little more than an outline of typical procedures of data collection for student evaluation. This is not the case. The first indication of a distinction between this and the usual testing model resides in steps two and three which assert that <u>all</u> determinants of success (within reason, of course) should be listed and assessed. The latter need not be as difficult as it may seem. The OSS reports,

> The multiform method of examination does not require ten or twelve times as many procedures as there are variables, because many procedures yield ratings on several different factors. Almost every factor, for instance, can be roughly estimated on the basis of an interview. Or, to take another example, one questionnaire can be constructed which lists every condition that is likely to be encountered in the field and which asks the subject to estimate the positive or negative appeal of each of these for him.[6]

Step four of "assessment" is sub-divided by the Assessment Staff in order to provide further guidance in the development of data-collection situations. First, the assessment program is to be embedded within a "social matrix composed of staff and candidates." Substituting "administrators, teachers, and peers" for "staff and candidates," we will assert the importance of observing students' behavior in terms of their usual or typical performance, as well as potential behavior, which we might assess in some other manner. Second, "several different types of procedures and several procedures of the same type for estimating the strength of each variable" should be developed. In the expanded evaluation model, we will want to be able to assess behaviors using bits of evidence of very different natures. The most reliable data will probably consist of standardized and teacher-made test scores. In addition, we will want to consider teacher ratings, language samples obtained in both natural and interview situations, a large number of distinct observations, reports of particular student difficulties or interests, reports of parent interest, and so on. In order to assure the reliability and validity of our assessments, we can make use of the principle of multiple measurements asserted here.

Consider a brief example. Objective 2 of the Pennsylvania Project asserts that students should develop an understanding and appreciation of different social, cultural, and ethnic groups. I'm sure that we can all envision situations in which success in meeting this objective is manifested. "Understanding" may be evaluated through traditional paper-and-pencil tests, for those students old enough to respond. Yet for the younger children or for the entire evaluation of "appre-

ciation," a number of unique bits of evidence may be necessary. We might wish to ask the child to discuss ethnic differences with us, we may wish to observe a child's behavior in an ethnically or racially integrated classroom, both during instruction and during "free" periods. Further, within a given free play period, two children may manifest "appreciation" in very different ways. While one young child may perhaps physically defend an ethnically different child, another may instead discuss the varied background with that individual. In addition, the manifestations of the attainment of a given objective will change with the individual's age.

Data collected under the principles discussed are not typical. For one, the data collection situations may be either of the usual standardized format or allowed to differ from student to student. The latter includes the option of utilizing observational data from naturally occurring situations. Second, the response required of the student to a given stimulus may not be of standardized format, and may even be "free-flowing" as would interview or conversation or free-play recordings. As a result, indicators of achievement may differ from student to student. This might occur also in courses which are of an individualized nature. A sixth-grade independent study course in science may be developing students' ability to investigate scientific phenomena with a minimum of teacher assistance. Such a class at first glance may appear completely without organization, having no more than two students pursuing a given acitivity simultaneously. Yet the more this disorganization may appear to prevail, the greater is the probability that the course objective is being met. Thus, the principles of assessment leave us with an extremely complex collection of relatively "unstructured data," and perhaps some question as to scoring and combining such data for specific evaluations.

Fortunately the problem is not without solution. Utilizing appropriate psychometric principles and with the assistance of technical tools as the digital computer, it is possible to obtain valid measures of behavior traits under the

conditions described. The process for doing so is one I have outlined but to date, have tested only in a single setting. Although the task in scoring complex data is of necessity technically complex, it is intuitively surprisingly simple, and with the aid of a computer technician could undoubtedly be simplified to the point where it could be applied by most school faculties.

The complete set of information obtained from any one subject or class comprises a single "behavior sample." The data may include test scores, observational data, reports and recordings of the behaviors themselves, such as interview transcripts or film recordings, teacher reports, and so on. To score the behavior samples for a given trait, set of traits, or course objective, a "behavior dictionary" is used as a referent. The behavior dictionary provides the answers to the question, "If the particular course objective is being met, how would this be manifested in the behavior samples collected?" The dictionary becomes a screen for selectively scoring bits of evidence contained in the behavior samples. The exact format of a behavior dictionary is not rigid. It might be, for example, a list of words which would occur in a language sample once certain vocabulary has been learned, or a set of categories of many behaviors which would be manifested by children who do (or do not) "appreciate different social, cultural, and ethnic groups." It might consist of a list of critical or cutoff scores for certain tests or frequencies of observed behaviors. In the one study conducted using this technique a word list was employed. In this case one question asked was, "What words or phrases might be used by the child in an unstructured discussion which would indicate a given level of egocentricity?" Obviously, the result of asking such a question could be an extremely long list of words and phrases, each perhaps accompanied by some weight according to its value as an indicator of the main variable, egocentricity. The job can be simplified, first by use of the computer for tasks of large-list handling, and second, by reference to the behavior samples themselves. That is, if no individual

under study manifests a particular behavior, or speaks a particular phrase, there is no need to include it in the behavior dictionary. In certain cases then the behavior dictionary may be constructed once some or all of the data have been collected.

A number of behavior dictionaries are already in existence. For behavior samples consisting of language recordings, there are the dictionaries which accompany the General Inquirer[7] and one developed by myself specifically for children's language. For evaluation of broader educational goals, as those of entire courses or course sequencies, we have the extensive "Cross Cultural Outline of Education," prepared by Anthropologist Jules Henry.[8] The Taxonomy of Educational Objectives[9] provides useful beginnings for behavior dictionaries involving specific sorts of educational outcomes. For extensive course evaluation however, appropriate behavior dictionaries are yet to be constructed. I will describe my current activities in doing so for certain primary school classes, in later paragraphs.

The data are projected onto one or more behavior dictionaries to yield the necessary final quantitative evidence. The derivation of quantitative partial indices of the major variables may take any of a number of forms and will vary from one evaluation to another. Thus, quantification techniques will not be discussed here except to note that ultimately most such techniques entail the identification, weighting, and summing of selected responses contained in a given behavior sample. The computer can be of obvious assistance by rapidly scanning large amounts of data for specific occurrences and trends. Generally multiple scores will result, and multivariate analysis techniques are useful for their summary. Extensive behavior samples may be re-analyzed a number of times to provide evidence on a number of major variables or course objectives.

The two-page handout which accompanies this paper outlines the expanded evaluation model and summarizes the basic principles on which it rests. We are now faced with the task of making the model operational in school settings. At the present

time I am undertaking this task for a variety of primary school situations. I have chosen to work at this level in particular for several reasons. One, a dearth of research, mostly recent, and supported to a great extent by Bloom's Stability and Change in Human Characteristics,[10] has indicated the extreme importance of the early schooling years as the time during which we have maximum chance to influence developing traits. In spite of this, the primary school remains the level of formal schooling for which there is the greatest paucity of operational class or school objectives, and of adequate evaluation devices. In addition we find a large number of educational innovations being used in the elementary grades, many of these too without sufficient evaluatory provisions.

The "job descriptions" for certain courses and grade levels are presently being formulated. To provide structure for the task, the model of the outcomes of instruction presented graphically in Figure 1 has been constructed. "Educational outcomes" are here viewed as the broadest range of changes which may occur in the behavior of students as a result of having been physically exposed to a formal educational environment and its accompanying characteristics (teacher or machine, materials, other pupils, etc.). Having realized the need for a model of evaluation to encompass both the effects of individual units of instruction and of larger units as several grade levels, or a sequence of courses in a given discipline, effects having been classified according to whether they are primarily associated with the former or the latter. The latter, effects upon the individual of a wide variety of learning experiences, have been termed "school" effects.[11] They would develop as the result of a quantity and variety of experiences in the educational setting, and involve a level of integration of specific learnings in the cognitive, affective, and/or habitual behavior areas. Frequently such general types of learnings are the "developmental tasks" necessary for proper socialization.

A second dimension on which educational outcomes may be classified involves general types of behavior which the individual may exhibit. The first two categories (cognitive - affective) have been borrowed from the "Taxonomies." In addition, an important class of behaviors not explicitly contained in these categories are the individuals' tendencies to make certain overt responses as a result of having internalized certain cognitive and affective abilities. The category has been named "habitual behavior" to indicate that it includes primarily what the person <u>does</u> typically do, as opposed to what he <u>may</u> potentially do. Thus observation of students in less structured situations will provide the majority of evidence on these behaviors.

I will not go into an extended discussion of the relationships, psychological or statistical, among the three categories of behavior, as they are obviously too numerous and too complex to relate in any detail. Cognitive and affective achievements are intimately related in the sense that strong positive affection for certain types of learning, or certain general personality traits, as McClelland's "need achievement"[12] will perhaps by definition, tend to compel the individual to seek additional learning. Bloom[13] proposes the exploitation of the reverse relationship, that is, increasing learning to a level of "mastery," in order to increase affect toward learning and toward a given discipline. "Habitual behaviors" are manifested as a result of the individual having attained given levels of cognitive and affective traits. Indeed, this third behavior category is so closely related to the others that we use it as a medium for their measurement. As a simple example, we frequently use voluntary reading as an indicator of interest in certain subject matter areas, or we may monitor a child's conversation to obtain a measure of the level of development of his grammar and usable vocabulary.

Figure 1 provides exemplary behaviors which are included in each of the six categories of educational outcomes. In order to complete the behavior dictionary

however, each category requires further definition. For example the class outcomes would be partitioned by course content. The measures of achievement in 1.a. would be subdivided according to, say, the levels of the cognitive Taxonomy, and the levels of 1.b. by its affective counterpart. Henry's "Cross-Cultural Outline" provides a basis for further partition of more general types of learning. The completed behavior dictionary is quite extensive and requires the use of the computer for efficient handling. Thus we are currently engaged in preparing programs for storing an extensive set of categories of behavior at a variety of levels, and for assigning weights and scores to the various categories.

The stated objectives of teachers and of school administrators form important subsets of the complete outline of outcomes. In order to identify these in particular, we are collecting the lists of objectives from the teachers and administrators, each major objective accompanied by a somewhat detailed operational definition. This will allow us to tag and score expected and unanticipated outcomes separately, and is also helpful in terms of identifying indicators of attainment of the various outcomes. These indicators include test scores, plus lists of other situations in which it is likely that the attainment of a given objective may be manifested in the individual's overt behavior. We find two additional resources of use in the task of identifying behaviors that provide "even a little evidence" on the attainment of the outcomes. The first of these is reference to past studies and outlines, such as the Taxonomy or the "Cross Cultural Outline." High on the list of reports of often unexpected and unevaluated instructional outcomes is Philip Jackson's interesting essay, Life in Classrooms.[14] Second, our own systematic observations of the classroom situation yield some unique and interesting behaviors to consider. For example, we have found that in a newly integrated second or third grade, Negro children are prone to raising their hands in response to questions, frequently without any knowledge of the correct answer. Although this in itself is a minor

observation, such behavior is a partial indicator of the level of success on other, more important outcomes of their instruction.

I would like to mention a set of habitual behaviors in which I have particular interest and which form an important part of our behavior dictionary. They involve the many effects of school upon language patterns of the pupils. Language development is an obvious and important "developmental task" for all of our pupils. Certain general developments of language behavior constitute some of the more important objectives of the entire primary school endeavor, and more specific developments and vocabulary, of secondary and higher education programs. A preliminary study I have conducted[15] indicates that in addition, there exist many high intercorrelations between connotative uses of language and other school achievements and developmental measures. For example correlations of .57 to .71 were found between the frequency of reference to one's self and other persons and a variety of intellectual measures including general intelligence and achievement in spelling, reading, and arithmetic; strong relationships were found between the use of affect-laden words, and age, social class, and various types of school achievement. The findings, which involve children from kindergarten through fourth grade, suggest ways in which specific language characteristics constitute an illuminating and important medium for the measurement of a wide variety of educational outcomes. Further, the study provides direction for means for the valid screening and scoring of trends in a particular behavior dictionary. (At least two other research teams[16] are currently employing similar types of language scores as evaluation media, one team reporting at this conference.)

When our list of indicators near completion (it will never be truly complete), it will be coded and stored for computer use with the behavior dictionary. Our

next steps will involve establishing rating, testing, and observational procedures for collecting the data necessary to screen with this behavior dictionary. In the coming academic year, we will be observing a large number of classrooms, and characterizing each class and school by its emphases in terms of the major instructional outcomes in the dictionary.

The system being proposed in this paper represents an attempt to extend our notions about evaluation to take into consideration current needs for comprehensive evaluations, for evaluation tools for new and changing curricula, and for the assessment of various groups of students who may be in some ways particularly difficult to test. In some ways, it is a proposal diametrically opposed to trends toward specificity and detailed analysis of formally stated educational outcomes. The extended evaluation model, which incorporates the principles of "assessment," is directed toward presenting a "total picture" of classroom and school effects upon the pupils. Admittedly data are difficult to analyze. In order to solve the former prcblem, assuming that we have a commitment to thorough evaluations, it may be necessary for us to adopt and extend Sorenson's definition of a formal role for the educational "evaluator."[17] Solution of the technical problems, while difficult, need only time, experience, and appropriate technical assistance from our colleagues. I am presently working toward overcoming these problems in order to make the extended evaluation model a usable and practical system for describing the variety of instructional outcomes realized by our students.

Footnotes

[1]Philip W. Jackson, Life in classrooms (New York: Holt, Rinehart and Winston, 1968), p. 123.

[2]Proceedings of the first meeting of the State Advisory Committee on the Assessment of Educational Quality (Harrisburg, Pennsylvania: Pennsylvania State Department of Public Instruction, February, 1968).

[3]Ralph W. Tyler, Needed research in mathematics education, Journal of Research and Development in Education, Fall, 1967, 1 (1), p. 140.

[4]The OSS Assessment Staff, Assessment of men (New York: Rinehart and Co., 1948).

[5]Since the original preparation of this paper, I have discovered with great interest that in response to similar needs and deficiencies, a similar set of evaluation principles has been outlined for use in the psychotherapeutic situation. These are reported in Robert I. Watson and Ivan N. Mensh, The evaluation of the effects of psychotherapy: 1. Sources of material, The Journal of Psychology, 1951, 32, 259-273. Also Watson, Mensh, and Edwin F. Glidea, same title: III. Research design, same Journal, 293-308.

[6]Assessment of men, p. 35.

[7]Philip J. Stone, et al. The General Inquirer (Cambridge, Mass.: M.I.T. Press, 1966).

[8]Jules Henry, A cross-cultural outline of education, Current Anthropologist, July, 1960, 1 (4), 267-305.

[9]Benjamin S. Bloom (Ed.), Taxonomy of educational objectives. Handbook I: Cognitive domain (New York: David McKay and Co., 1956). David R. Krathwohl, Benjamin S. Bloom, and Bertram B. Masia, Taxonomy of educational objectives. Handbook II: Affective domain (New York: David McKay and Co., 1964).

[10]Benjamin S. Bloom, Stability and change in human characteristics (New York: John Wiley and Sons, 1964).

[11]The classification of educational outcomes on this dimension is suggested by William A. Burton in The guidance of learning activities (3d. ed.; New York: Appleton-Century-Crofts, 1962).

[12]D. C. McClelland, J. W. Atkinson, and E. L. Lowell, The achievement motive (New York: Appleton-Century-Crofts, 1953).

[13]Benjamin S. Bloom, Learning for mastery, UCLA Center for the Study of Evaluation in Instructional Programs Evaluation Comment, May, 1968, 1 (2), 1-12.

Footnotes
(Continued)

[14]Jackson, loc. cit.

[15]Jeremy D. Finn, Patterns in children's language, The School Review (in press).

[16]Clarence H. Bradford, The Inquirer II and content analysis used for evaluation of a special program for slow learners. Paper to be read at the 1969 Annual Meeting of the American Educational Research Association, Los Angeles, February 5-8, session 10.14. Also William E. Gardner and Paul E. Johnson, Evaluation of language habits in a behavioral science curriculum, The School Review, December, 1968, 76 (4), 396-411.

[17]Garth Sorenson, A new role in education: The evaluator, UCLA Center for the Study of Evaluation in Instructional Programs Evaluation Comment, January, 1968, 1 (1), 1-4.

EVALUATION OF INSTRUCTIONAL OUTCOMES: THE USE OF UNSTRUCTURED DATA

Jeremy D. Finn
Faculty of Educational Studies
State University of New York at Buffalo

Eugene R. Smith, Ralph W. Tyler, and the "Evaluation Staff" in Appraising and Recording Student Progress, list these "basic assumptions of evaluation":

1. Education is a process which seeks to change behavior patterns of human beings.
2. The kinds of changes in behavior patterns in human beings which the school seeks to bring about are the educational objectives.
3. An educational program is appraised by finding out how far the objectives of the program are actually being met.
4. Human behavior is ordinarily so complex that it cannot be adequately described or measured by a single term or single dimension.
5. The way in which a student organizes his behavior patterns is an important aspect to be appraised.
6. The methods of evaluation are not limited to the giving of paper-and-pencil tests; any device which provides valid evidence regarding the progress of students toward educational objectives is appropriate.
7. The nature of the appraisal influences teaching and learning.
8. Responsibility for evaluating the school program belong(s) to the staff and clientele of the school.

The present paper suggests that although the primary purpose of evaluation is to determine the extent to which the objectives are being realized, this alone is not sufficiently comprehensive for current needs. A number of outcomes of instruction not intended, as well as those expected, need to be described. In order to do so, a system of assessment is recommended which stresses the application of principle 4 concerning the collection of a variety of evaluatory measures.

Unique characteristics of the expanded model

1. Educational "achievement" is viewed as involving numerous types of behavior changes. As such, the outcome of the system is a holistic view of each individual or class.
2. A framework is provided for obtaining evaluatory data on special groups of students who may not be amenable to testing by usual paper-and-pencil instruments. Also allows for the evaluation of a variety of non-cognitive behaviors, e.g. attitudes, habits, activities, personality and language variables.
3. Evaluations may differ in stimulus and response from one individual to another. Two individuals may have comparable overall levels of a given trait, but may manifest them quite differently. This facility provides a partial solution to problems of non-comparability of test results for various experimental groups or for subjects across grade levels or schools.
4. Behavior samples may be re-analyzed at later points in time to obtain information on other variables of interest or on the progress of certain behaviors over the grades.
5. Expensive in time, money, and facilities to administer. To begin the application of the full model may require the services of a professional evaluator, in addition to specialists in measurement and computer technology.

Below are listed the general evaluation procedures as outlined by Smith and Tyler. The right-hand column lists the modifications in procedure necessary to combine the assessment procedures with it.

Operational Procedures	
Original Evaluation Model	Extended Model
1. Formulate objectives	1. List potential instructional outcomes a. Teachers' objectives b. School and societal objectives c. Unspecified potential outcomes derived from observation and prior study of classes of interest.
2. Classify objectives	
3. Define objectives in terms of pupil behavior	2. Formulate behavior dictionary; list the many ways in which achievement of each outcome listed may be shown.* Assign weights to each according to its importance as an indicator of achievement of each outcome listed.
4. Suggest situations in which achievement of the objectives will be shown.	
5. Select and try promising evaluation methods	3. Construct tests, define rating scales, etc. Collect behavior samples.
6. Develop and improve appraisal methods	4. Select and score items from behavior samples, using measurement and computer technological assistance. Adjust relative weights of various bits of evidence.
7. Interpret results	5. Write complete description of achievement of each individual or class. Hold staff conference for reviewing and for suggesting student placement and curriculum modifications.

* Consider the types of data available, i.e. cognitive-affective testing results, observations of academic or free time performance, teacher ratings, pupil self-ratings, structured or unstructured interview or language data, observations or reports of out-of-class activities. The behavior dictionary may be modified according to the particular behavior samples collected, to allow each student to contribute indicators of achievement not manifested by his peers.

	Class (a)	School (b)
Cognitive Behavior (1)	1.a. Knowledge of facts in content areas Comprehension, translation, interpretation, extrapolation, application, analysis, evaluation, in specific courses Potential vocabulary	1.b. Synthesis of facts within and between course areas Organization of cognitive behaviors Knowledge of scientific process, generalizations concerning human, animal and plant behavior, science and mathematics Can apply knowledge in real-life situations Means of obtaining and maintaining employment (See Henry's "Cross-Cultural Outline")
Affective Behavior (2)	2.a. Attitudes and interests in specific subject-matter areas Attitude toward specific events, e.g. recent scientific accomplishments, political events	2.b. Attitudes toward school, learning, cultures and ethnic groups, toward the disciplines, reading Self-image and "needs" to achieve, to be liked, etc. Political attitudes Attitudes toward individual differences Attitude toward religion, art
Habitual Behavior (3)	3.a. Reads in subject matter area Is precise in his thinking Uses the learned vocabulary Uses the principles of certain courses (e.g. arithmetic) in daily life	3.b. Reads Studies Seeks new knowledge Defends individual differences Participates in community activity Maintains own health Attends lectures, art demonstrations, etc. Is critical in appropriate places Supports other individuals

Figure 1 The classification of the effects of formal educational processes